run_analysis.R

jamesmckay

2020-08-20

```r
# Getting and Cleaning Data Project



#Load Packages
library(data.table)
library(tidyverse)

## ── Attaching packages ─────────────────────────────────
tidyverse 1.3.0 ──

##   ggplot2 3.3.2      purrr   0.3.4
##   tibble  3.0.3      dplyr   1.0.0
##   tidyr   1.1.0      stringr 1.4.0
##   readr   1.3.1      forcats 0.5.0

## ── Conflicts ──────────────────────────────────
tidyverse_conflicts() ──
## x dplyr::between()   masks data.table::between()
## x dplyr::filter()    masks stats::filter()
## x dplyr::first()     masks data.table::first()
## x dplyr::lag()       masks stats::lag()
## x dplyr::last()      masks data.table::last()
## x purrr::transpose() masks data.table::transpose()

library(dplyr)



# Check if directory exists if not create data directory and download
file
if (!file.exists("./data")){dir.create("./data")
  fileURL <-
"https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20H
AR%20Dataset.zip"
  download.file(fileURL, filename, method="curl")
  }



# Unzip folder
unzip(zipfile = "./data/UCI HAR Dataset")
```

```
## Warning in unzip(zipfile = "./data/UCI HAR Dataset"): error 1 in
extracting from
## zip file

list.files("UCI HAR Dataset")

## [1] "activity_labels.txt" "features_info.txt"   "features.txt"
## [4] "README.txt"          "test"                "train"

# Read in all the data sets into environment

# read in features txt.file
features <- read.table("UCI HAR Dataset/features.txt", col.names =
c("n","functions"))

#read in activities txt.file
activities <- read.table("UCI HAR Dataset/activity_labels.txt",
col.names = c("code", "activity"))

#read in test txt.files
subject_test <- read.table("UCI HAR Dataset/test/subject_test.txt",
col.names = "subject")
x_test <- read.table("UCI HAR Dataset/test/X_test.txt", col.names =
features$functions)
y_test <- read.table("UCI HAR Dataset/test/y_test.txt", col.names =
"code")

#read in train txt.files
subject_train <- read.table("UCI HAR Dataset/train/subject_train.txt",
col.names = "subject")
x_train <- read.table("UCI HAR Dataset/train/X_train.txt", col.names =
features$functions)
y_train <- read.table("UCI HAR Dataset/train/y_train.txt", col.names =
"code")

#1: Merges the training and the test sets to create one data set.
x_merge <- rbind(x_train, x_test)
y_merge <- rbind(y_train, y_test)
subject_merge <- rbind(subject_train, subject_test)
merge_data  <- cbind(subject_merge, y_merge, x_merge)

#2: Extracts only the measurements on the mean and standard deviation
for each measurement.
tidy_data <- merge_data %>% select(subject, code, contains("mean"),
contains("std"))

#step3: Uses descriptive activity names to name the activities in the
data set.
tidy_data$code <- activities[tidy_data$code, 2]
```

```r
#step4: Appropriately labels the data set with descriptive variable
names.

names(tidy_data)[2] = "activity"
names(tidy_data)<-gsub("Acc", " Accelerometer", names(tidy_data))
names(tidy_data)<-gsub("Gyro", " Gyroscope", names(tidy_data))
names(tidy_data)<-gsub("BodyBody", " Body", names(tidy_data))
names(tidy_data)<-gsub("Mag", " Magnitude", names(tidy_data))
names(tidy_data)<-gsub("^t", " Time", names(tidy_data))
names(tidy_data)<-gsub("^f", " Frequency", names(tidy_data))
names(tidy_data)<-gsub("tBody", " TimeBody", names(tidy_data))
names(tidy_data)<-gsub("-mean()", " Mean", names(tidy_data),
ignore.case = TRUE)
names(tidy_data)<-gsub("-std()", " STD", names(tidy_data), ignore.case
= TRUE)
names(tidy_data)<-gsub("-freq()", " Frequency", names(tidy_data),
ignore.case = TRUE)
names(tidy_data)<-gsub("angle", " Angle", names(tidy_data))
names(tidy_data)<-gsub("gravity", " Gravity", names(tidy_data))

#step5: From the data set in step 4, creates a second, independent tidy
data set with the average of each variable for each activity and each
subject.

summary_data <- tidy_data %>%
  group_by(subject, activity) %>%
  summarise_all(funs(mean))

## Warning: `funs()` is deprecated as of dplyr 0.8.0.
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was
generated.

write.table(summary_data, "summary_data.txt", row.name=FALSE)

summary_data

## # A tibble: 180 x 88
## # Groups:   subject [30]
##     subject activity ` TimeBody Acce… ` TimeBody Acce… ` TimeBody
```

```
Acce…
##        <int> <fct>                <dbl>            <dbl>
<dbl>
##  1        1 LAYING               0.222          -0.0405              -
0.113
##  2        1 SITTING              0.261          -0.00131             -
0.105
##  3        1 STANDING             0.279          -0.0161              -
0.111
##  4        1 WALKING              0.277          -0.0174              -
0.111
##  5        1 WALKING…             0.289          -0.00992             -
0.108
##  6        1 WALKING…             0.255          -0.0240              -
0.0973
##  7        2 LAYING               0.281          -0.0182              -
0.107
##  8        2 SITTING              0.277          -0.0157              -
0.109
##  9        2 STANDING             0.278          -0.0184              -
0.106
## 10        2 WALKING              0.276          -0.0186              -
0.106
## # … with 170 more rows, and 83 more variables: `TimeGravity
## #   Accelerometer.mean...X` <dbl>, `TimeGravity
Accelerometer.mean...Y` <dbl>,
## #   `TimeGravity Accelerometer.mean...Z` <dbl>, `TimeBody
## #   AccelerometerJerk.mean...X` <dbl>, `TimeBody
## #   AccelerometerJerk.mean...Y` <dbl>, `TimeBody
## #   AccelerometerJerk.mean...Z` <dbl>, `TimeBody
Gyroscope.mean...X` <dbl>, `
## #   TimeBody Gyroscope.mean...Y` <dbl>, `TimeBody
Gyroscope.mean...Z` <dbl>, `
## #   TimeBody GyroscopeJerk.mean...X` <dbl>, `TimeBody
## #   GyroscopeJerk.mean...Y` <dbl>, `TimeBody
GyroscopeJerk.mean...Z` <dbl>, `
## #   TimeBody Accelerometer Magnitude.mean..` <dbl>, `TimeGravity
Accelerometer
## #   Magnitude.mean..` <dbl>, `TimeBody AccelerometerJerk
## #   Magnitude.mean..` <dbl>, `TimeBody Gyroscope Magnitude.mean..`
<dbl>, `
## #   TimeBody GyroscopeJerk Magnitude.mean..` <dbl>, `FrequencyBody
## #   Accelerometer.mean...X` <dbl>, `FrequencyBody
## #   Accelerometer.mean...Y` <dbl>, `FrequencyBody
## #   Accelerometer.mean...Z` <dbl>, `FrequencyBody
## #   Accelerometer.meanFreq...X` <dbl>, `FrequencyBody
## #   Accelerometer.meanFreq...Y` <dbl>, `FrequencyBody
## #   Accelerometer.meanFreq...Z` <dbl>, `FrequencyBody
## #   AccelerometerJerk.mean...X` <dbl>, `FrequencyBody
## #   AccelerometerJerk.mean...Y` <dbl>, `FrequencyBody
```

```
## #    AccelerometerJerk.mean...Z` <dbl>, ` FrequencyBody
## #    AccelerometerJerk.meanFreq...X` <dbl>, ` FrequencyBody
## #    AccelerometerJerk.meanFreq...Y` <dbl>, ` FrequencyBody
## #    AccelerometerJerk.meanFreq...Z` <dbl>, ` FrequencyBody
## #    Gyroscope.mean...X` <dbl>, ` FrequencyBody Gyroscope.mean...Y`
<dbl>, `
## #    FrequencyBody Gyroscope.mean...Z` <dbl>, ` FrequencyBody
## #    Gyroscope.meanFreq...X` <dbl>, ` FrequencyBody
## #    Gyroscope.meanFreq...Y` <dbl>, ` FrequencyBody
## #    Gyroscope.meanFreq...Z` <dbl>, ` FrequencyBody Accelerometer
## #    Magnitude.mean..` <dbl>, ` FrequencyBody Accelerometer
## #    Magnitude.meanFreq..` <dbl>, ` Frequency Body AccelerometerJerk
## #    Magnitude.mean..` <dbl>, ` Frequency Body AccelerometerJerk
## #    Magnitude.meanFreq..` <dbl>, ` Frequency Body Gyroscope
## #    Magnitude.mean..` <dbl>, ` Frequency Body Gyroscope
## #    Magnitude.meanFreq..` <dbl>, ` Frequency Body GyroscopeJerk
## #    Magnitude.mean..` <dbl>, ` Frequency Body GyroscopeJerk
## #    Magnitude.meanFreq..` <dbl>, ` Angle. TimeBody
AccelerometerMean.
## #    Gravity.` <dbl>, ` Angle. TimeBody AccelerometerJerkMean..
## #    GravityMean.` <dbl>, ` Angle. TimeBody GyroscopeMean.
GravityMean.` <dbl>,
## #    ` Angle. TimeBody GyroscopeJerkMean. GravityMean.` <dbl>, `
Angle.X.
## #    GravityMean.` <dbl>, ` Angle.Y. GravityMean.` <dbl>, ` Angle.Z.
## #    GravityMean.` <dbl>, ` TimeBody Accelerometer.std...X` <dbl>, `
TimeBody
## #    Accelerometer.std...Y` <dbl>, ` TimeBody Accelerometer.std...Z`
<dbl>, `
## #    TimeGravity Accelerometer.std...X` <dbl>, ` TimeGravity
## #    Accelerometer.std...Y` <dbl>, ` TimeGravity
Accelerometer.std...Z` <dbl>, `
## #    TimeBody AccelerometerJerk.std...X` <dbl>, ` TimeBody
## #    AccelerometerJerk.std...Y` <dbl>, ` TimeBody
## #    AccelerometerJerk.std...Z` <dbl>, ` TimeBody Gyroscope.std...X`
<dbl>, `
## #    TimeBody Gyroscope.std...Y` <dbl>, ` TimeBody Gyroscope.std...Z`
<dbl>, `
## #    TimeBody GyroscopeJerk.std...X` <dbl>, ` TimeBody
## #    GyroscopeJerk.std...Y` <dbl>, ` TimeBody GyroscopeJerk.std...Z`
<dbl>, `
## #    TimeBody Accelerometer Magnitude.std..` <dbl>, ` TimeGravity
Accelerometer
## #    Magnitude.std..` <dbl>, ` TimeBody AccelerometerJerk
## #    Magnitude.std..` <dbl>, ` TimeBody Gyroscope Magnitude.std..`
<dbl>, `
## #    TimeBody GyroscopeJerk Magnitude.std..` <dbl>, ` FrequencyBody
## #    Accelerometer.std...X` <dbl>, ` FrequencyBody
Accelerometer.std...Y` <dbl>,
## #    ` FrequencyBody Accelerometer.std...Z` <dbl>, ` FrequencyBody
```

```
## #   AccelerometerJerk.std...X` <dbl>, ` FrequencyBody
## #   AccelerometerJerk.std...Y` <dbl>, ` FrequencyBody
## #   AccelerometerJerk.std...Z` <dbl>, ` FrequencyBody
Gyroscope.std...X` <dbl>,
## #   ` FrequencyBody Gyroscope.std...Y` <dbl>, ` FrequencyBody
## #   Gyroscope.std...Z` <dbl>, ` FrequencyBody Accelerometer
## #   Magnitude.std..` <dbl>, ` Frequency Body AccelerometerJerk
## #   Magnitude.std..` <dbl>, ` Frequency Body Gyroscope
Magnitude.std..` <dbl>,
## #   ` Frequency Body GyroscopeJerk Magnitude.std..` <dbl>
```